

Compute resources available to the research community for microbiome analysis

Bhavya Papudeshi, Sheri Sanders, Carrie Ganote, Thomas G. Doak



[Home](#)[Training](#)[NCGAS Staff](#)[Get Help](#)

NCGAS is an NSF-funded center tasked with assisting biologists in gaining access to computational resources needed to analyze genomic data (it's free!). This is done by providing:

- Access to compute resources to run their analyses
- Bioinformatics support- select pipelines/programs that best analyze the data to address research questions
- Training through workshops, sharing developed pipelines, online courses, and materials
- Gateways like Galaxy, GenePattern and genome browsers

Why should you care?

- Attending a bioinformatics workshop that has asked you to install a couple of programs beforehand



Copyright 2004 by Randy Glasbergen.
www.glasbergen.com



**"This software will help you manage stress
as long as you don't try to install it."**

https://www.glasbergen.com/ngg_tag/manage-stress/

Jetstream, a cloud computing environment

- Jetstream is an XSEDE (The Extreme Science and Engineering Discovery Environment) resource. XSEDE is an NSF project that makes available cyberinfrastructure to share data, run analysis and provides the expertise.



- Like cloud computing environments, you can spin up a virtual machine.



Google Cloud

Jetstream, a cloud computing environment

- Virtual Machines are similar to a laptop, but computation is done elsewhere
- Users have root privileges, unlike high performance clusters
- Virtual machines help with the transition to command line use, software installation, and running analysis in the Linux environment
- Develop workflows or spin up pre-configured virtual machines, which can be shared and published between collaborators or publicly



Indiana University and Texas Advanced
Computing Center (TACC) host the XSEDE
cloud environment called Jetstream



Jetstream

NCGAS use cases for Jetstream

- Teaching workshops, was able to give all 50 workshop participants the exact same R setup, including all packages they needed.
- Used to provision a bioinformatics course at Bethane Cookman University.
- A prebuilt genome browser that eliminates much of the command line necessary to get started, to visualize data you just have to upload the data.
- MATLAB at no cost.
- Genomics toolkit image available with more than 30 packages already installed.
- Specialized bioinformatics pipelines- such QIIME2, PacBio, Galaxy service also available.

Getting access to Jetstream

Step 1- Getting an XSEDE account

First –an XSEDE account, create one at <https://portal.xsede.org>

The screenshot shows the XSEDE User Portal homepage. The header includes the XSEDE logo, the text "Extreme Science and Engineering Discovery Environment", a search bar, and a "SIGN IN" button. Below the header is a navigation bar with links: MY XSEDE, RESOURCES, DOCUMENTATION, ALLOCATIONS, TRAINING, USER FORUMS, HELP, ECSS, and ABOUT. Under "MY XSEDE" are links for Summary, Allocations/Usage, Accounts, Jobs, Profile, Publications, Tickets, Change Password, Add User, Community Accounts, and SSH Terminal.

Get Started on XSEDE

- Sign In
- Create Account

Quick Links

- System Monitor
- Allocations
- User News

In The Past 7 Days

XD SUs Charged: Total: by Field of Science

| Field of Science | Value |
|---|--------------|
| Materials Research | 10,032,719.0 |
| Biophysics | 6,183,797.0 |
| Chemistry | 4,905,945.0 |
| Condensed Matter Physics | 4,351,681.0 |
| Physics | 2,195,108.0 |
| Biological Sciences | 2,192,548.0 |
| Biochemistry and Molecular Structure and Function | 2,071,395.0 |
| Molecular Biosciences | 1,824,166.0 |
| Systematic and Population Biology | 1,744,875.0 |
| Integrative Biology and Neuroscience | 1,697,746.0 |
| All 62 others | 18,344,328.0 |

Number of Jobs Charged by Field of Science

Number of Jobs Charged by Field of Science

View Gallery

2017-03-31 16:00:00 UTC. Data: 2017-03-31 16:00:00 UTC. Powered by: XSEDE/High-Res

Jetstream, a cloud computing environment

Go to **www.jetstream-cloud.org**

Apps SDSU NCGAS Hacks Conferences JS Papers to read Bacteriocins Executing PhyloSift



Jetstream login

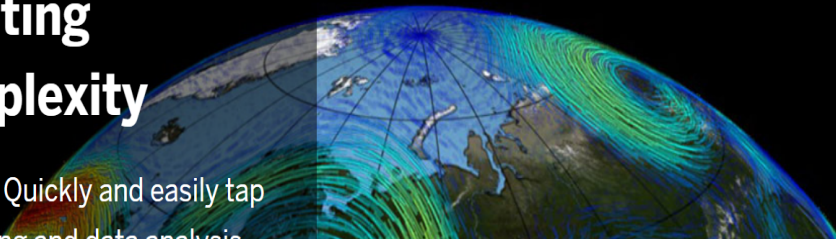
Request account

About Support Tech specs News & events Research Contact us



Research computing without the complexity

Jetstream is very user-friendly. Quickly and easily tap
into high performance computing and data analysis



Step 2- Selecting a Jetstream allocation

First you will need a Jetstream allocation to get started,

1. Request for a Jetstream trial access- limited resources <http://wiki.jetstream-cloud.org/Jetstream+Trial+Access+Allocation>
2. Contact NCGAS (help@ncgas.org) to get on our allocation – trial access as well but more resources than Option 1
3. Request for your own Jetstream allocation- <http://wiki.jetstream-cloud.org/Jetstream+Allocations>

Preconfigured biology based VM's on Jetstream

← → ↻ Secure | <https://use.jetstream-cloud.org/application/images> 🔍 ☆





Apps SDSU NCGAS Hacks Conferences JS Papers to read Bacteriocins Executing PhyloSift |

Jetstream Images ? Help Login

bioinfo

Showing 22 results for 'bioinfo'

All Images

| | | |
|---|--|---|
|  | BioLinux 8 Dec 5th 17 09:04 by jfischer | Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg -- **REQUIRES m1.small instance ... bioinformatics desktop Featured gui m1_small Ubuntu x2go |
|  | SearchSRA Base Image Nov 16th 17 04:31 by linsal | A base installation of the tools to download sequences and search the Sequence Read Archiv ... bioinformatics samtools SDSU SRA |
|  | bgRAMOSE - Genomics Software Nov 10th 17 12:15 by vbrendel | This image provides a platform for Brendel Group RAMOSE workflows for computational genom ... bioinformatics genomics gui NGS Ubuntu |
|  | CAFE Phylogenetics Tutorial | CAFE (Computational Analysis of gene Family Evolution) is an application that analyzes cha ... |

There are more than 20 preconfigured bioinformatics related images currently hosted on Jetstream.

Configuring a Virtual Machine on Jetstream

Don't see the software or workflow necessary for your research

Login ➡ startup a Ubuntu/CentOS image ➡ install software ➡ save/publish

- Install all the tools necessary for the workflow to the virtual machines
- Click on Image-> add image name and description, tags -> Next
- Type in the version number, base image used -> Next
- Access: public, specific users or private -> Submit

Generates a ticket to help@xsede.edu and wait for Jetstream team to image it for you.

Microbiome tools on preconfigured virtual machines

Microbiome related Jetstream images

- **Anvi'o** - Anvi'o is an open-source, community-driven analysis and visualization platform for 'omics data. [Link here](#)
- **QIIME2 toolkit** - Quantitative Insights into Microbial Ecology is an open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data. [Link here](#)
- **MEGAN** - "MEtaGenome ANalyzer" is a computer program that allows optimized analysis of large metagenomic datasets. [Link here](#)
- **Rstudio for NCGAS**- Set up and installed packages for various NCGAS workshops. Includes many genomic and mapping R packages and preloads three default users. [Link here](#)
- **PacBio SMRT Analysis** – PacBio analysis tools. [Link here](#)

There are more than 20 bioinformatics virtual machines

Start a preconfigured VM- Anvi'o

56% of 200000 SUs from TG-MCB170036

Instance Details

| | |
|------------|---|
| Status | ● Active |
| Activity | N/A |
| Size | m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB) |
| IP Address | 149.165.156.231 Copy |
| Launched | Oct 13, 2017 (3 days ago) |
| Based on | Ubuntu 16.04 Devel and Docker v.1.7 v.1.7 |
| Provider | Jetstream - Indiana University |
| ID | 17614 |
| Alias | 261988e0-919c-43ee-b53c-005ae354dd68 Copy |

■ Stop

🔄 Reboot

🔄 Redeploy

✖ Delete

Links

🔗 Open Web Shell [🔗](#)

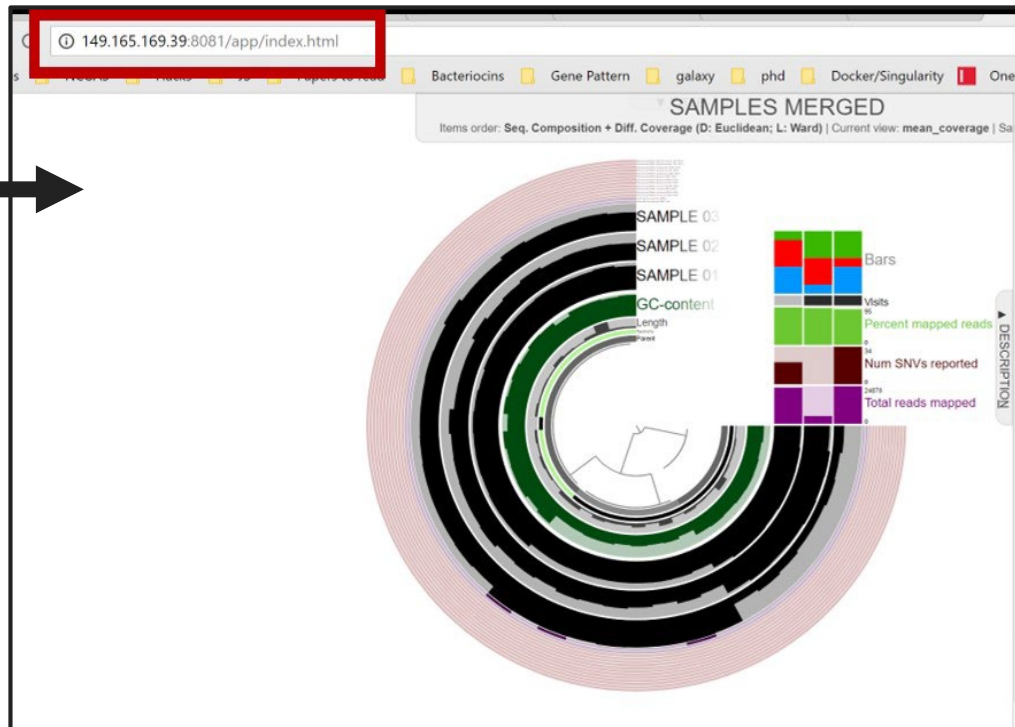
🖥️ Open Web Desktop [🔗](#)

🖥️ Open New Web Shell (beta) [🔗](#)

🖥️ Open New Web Desktop
(beta) [🔗](#)

Start a preconfigured VM- Anvi'o

```
Applications GhostKOALA - Job Req... bhnala@js-169-39: ~  
bhnala@js-169-39: ~  
File Edit View Search Terminal Help  
bhnala@js-169-39:~$ anvi-profile --version  
Anvi'o version .....: margaret (vunknown)  
Profile DB version .....: 29  
Contigs DB version .....: 12  
Pan DB version .....: 12  
Genome data storage version .....: 6  
Auxiliary data storage version .....: 2  
Structure DB version .....: 1  
bhnala@js-169-39:~$ anvi-self-test --suite mini
```



For more information -
[Running Anvio on Jetstream](#)

Acknowledgments



NCGAS Jetstream blog –

[Getting Started on Jetstream](#)

[Setting up SSH keys](#)

[Setting up Globus on Jetstream](#)

[Setting up Volumes on Jetstream](#)

[Adding users to Jetstream image](#)

[Using Ansible to manage Jetstream images](#)

[Running Anvio on Jetstream](#)

[NCGAS maintained Jetstream images](#)

Jetstream documentation - <https://iujetstream.atlassian.net/wiki/spaces/JWT/overview>

Contact NCGAS- help@ncgas.org



@ncgasiu



@ncgas



